

# Nobel Prize Winners

Sharon Morris

1/13/2019

## 1. The most Nobel of Prizes

The Nobel Prize is perhaps the worlds most well known scientific award. Except for the honor, prestige and substantial prize money the recipient also gets a gold medal showing Alfred Nobel (1833 - 1896) who established the prize. Every year it's given to scientists and scholars in the categories chemistry, literature, physics, physiology or medicine, economics, and peace. The first Nobel Prize was handed out in 1901, and at that time the Prize was very Eurocentric and male-focused, but nowadays it's not biased in any way whatsoever. Surely. Right?

The Nobel Foundation has made a dataset available of all prize winners from the start of the prize, in 1901, to 2016. Let's load it in and take a look.

```
library(tidyverse)
library(knitr)
library(lubridate)

nobel <- read_csv('https://raw.githubusercontent.com/indianspice/Data-Manipulation/master/Nobel%20Prize.csv')

kable(head(nobel), format = "markdown", padding = 0)
```

year	category	prize	motivation	prize amount	date of birth	full name	birth date	birth place	birth country	education	organization	aid amount	aid date	aid country	
1901	Chemistry	First Nobel Prize in Chemistry 1901	“in recognition of the extraordinary services he has rendered by the discovery of the laws of chemical dynamics and osmotic pressure in solutions”	1/1 160 000	Indiv	Janus Henricus van 't Hoff	1852-08-30	Rotterdam	Netherlands	University of Berlin	Germany	1901-03-01	1901-03-01	Germany	
1901	Literature	First Nobel Prize in Literature 1901	“in special recognition of his poetic composition, which gives evidence of lofty idealism, artistic perfection and a rare combination of the qualities of both heart and intellect”	1/1 569 000	Indiv	Sully Prudhomme	1839-03-16	Paris	France	NA	NA	NA	1901-09-07	1901-09-07	France

year	category	prize	motivation	prize_share	prize_full	type	birth_year	birth_city	birth_country	sex	organization	organization_country	award_year	award_city	award_country
1901	Medicine	The Nobel Prize in Physiology or Medicine	“for his work on serum therapy, especially its application against diphtheria, by which he has opened a new road in the domain of medical science and thereby placed in the hands of the physician a victorious weapon against illness and deaths”	1/1	293	Individual	1854	Hans Adolf Behring	Prussia (Germany)	Male	Marburg University	Germany	1901	Marburg	Germany
1901	Peace	The Nobel Peace Prize	NA	1/2	462	Individual	1828	Henry Dunant	Geneva (Switzerland)	Male	NA	NA	1901	Heidelberg	Switzerland
1901	Peace	The Nobel Peace Prize	NA	1/2	463	Individual	1822	Fledder Passy	Paris (France)	Male	NA	NA	1901	Paris	France
1901	Physics	The Nobel Prize in Physics	“in recognition of the extraordinary services he has rendered by the discovery of the remarkable rays subsequently named after him”	1/1	1	Individual	1845	Wilhelm Conrad Röntgen	Prussia (Germany)	Male	Munich University	Germany	1901	Munich	Germany

## 2. So, who gets the Nobel Prize?

All of the winners in 1901 were guys that came from Europe. But that was back in 1901, looking at all winners in the dataset, from 1901 to 2016, which sex and which country is the most commonly represented?

(For country, we will use the birth\_country of the winner, as the organization\_country is NA for all shared Nobel Prizes.)

```
# Counting the number of (possibly shared) Nobel Prizes handed
# out between 1901 and 2016
kable(nobel %>%
  count(prize_share), caption = "Shared Nobel Prizes")
```

Table 2: Shared Nobel Prizes

prize_share	n
1/1	344
1/2	306
1/3	201
1/4	60

*# Counting the number of prizes won by male and female recipients.*

```
kable(nobel %>%
  count(sex) %>%
  group_by(sex))
```

sex	n
Female	49
Male	836
NA	26

*# Counting the number of prizes won by different nationalities.*

```
kable(nobel %>%
  count(birth_country) %>%
  group_by(birth_country) %>%
  arrange(desc(n)) %>%
  head(n=20), format = "markdown")
```

birth_country	n
United States of America	259
United Kingdom	85
Germany	61
France	51
Sweden	29
NA	26
Japan	24
Canada	18
Netherlands	18
Italy	17
Russia	17
Switzerland	16
Austria	14
Norway	12
China	11
Denmark	11
Australia	10
Belgium	9
Scotland	9
South Africa	9

### 3. USA dominance

The most common Nobel laureate between 1901 and 2016 was a man born in the United States of America. But in 1901 all the laureates were European. When did the USA start to dominate the Nobel Prize charts?

*# Calculating the proportion of USA born winners per decade*

```
kable(prop_usa_winners <- nobel %>%
  mutate(usa_born_winner = birth_country == "United States of America",
    decade = year - (year %>% 10)) %>%
  group_by(decade) %>%
  summarize(proportion = mean(usa_born_winner, na.rm = TRUE)) %>%
```

```
print(prop_usa_winner))
```

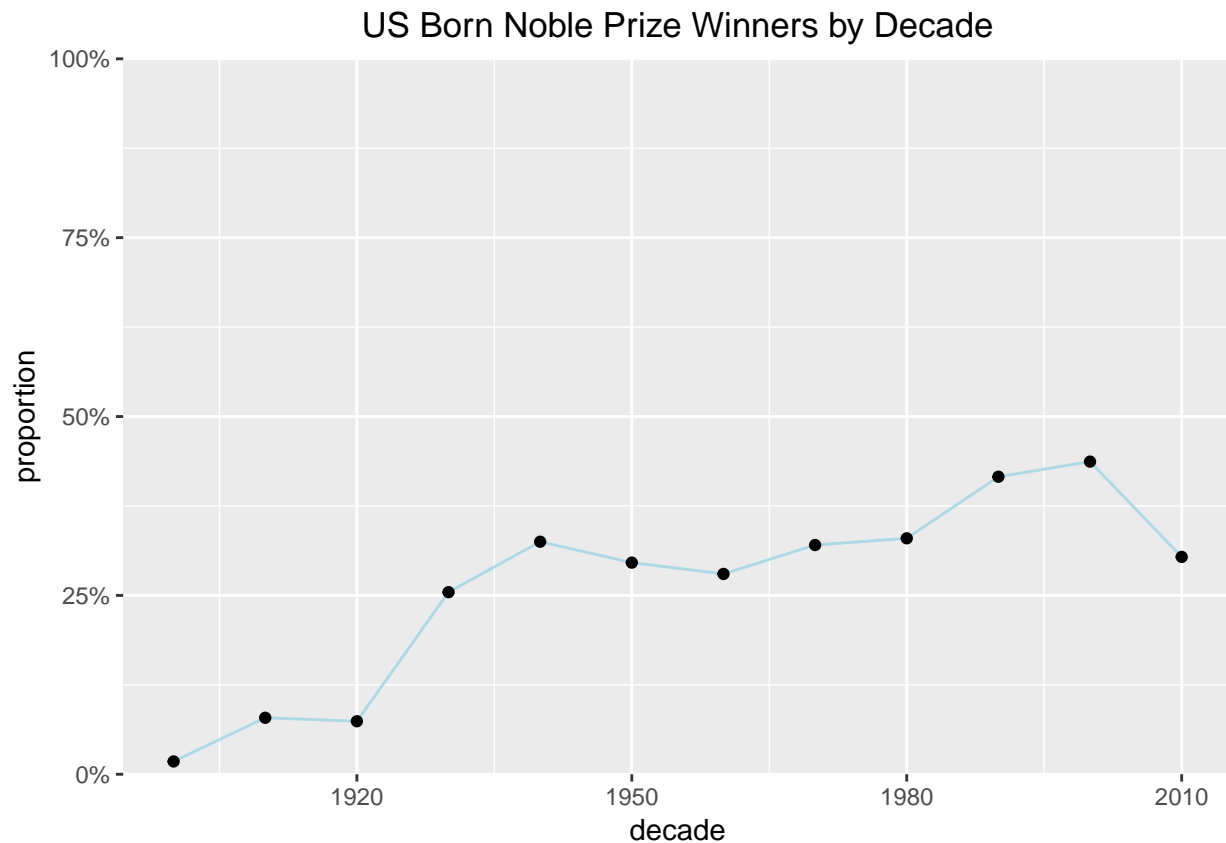
```
## # A tibble: 12 x 2
##   decade proportion
##   <dbl>      <dbl>
## 1  1900      0.0179
## 2  1910      0.0789
## 3  1920      0.0741
## 4  1930      0.255
## 5  1940      0.325
## 6  1950      0.296
## 7  1960      0.28
## 8  1970      0.320
## 9  1980      0.330
## 10 1990      0.416
## 11 2000      0.437
## 12 2010      0.304
```

```
decade proportion ——— 1900 0.0178571 1910 0.0789474 1920 0.0740741 1930 0.2545455 1940
0.3250000 1950 0.2957746 1960 0.2800000 1970 0.3203883 1980 0.3297872 1990 0.4158416 2000 0.4369748
2010 0.3037975 ## 4. USA dominance, visualized
```

A table is OK, but to see when the USA started to dominate the Nobel charts we need a plot!

```
# Setting the size of plots in this notebook
options(repr.plot.width=7, repr.plot.height=4)

ggplot(prop_usa_winners, aes(x = decade, y = proportion)) +
  geom_line(color = "light blue") +
  geom_point() +
  scale_y_continuous(labels = scales::percent,
                     limits = 0:1, expand = c(0,0)) +
  ggtitle("US Born Noble Prize Winners by Decade") +
  theme(plot.title = element_text(hjust = 0.5))
```



## 5. What is the gender of a typical Nobel Prize winner?

The USA became the dominating winner of the Nobel Prize first in the 1930s and has kept the leading position ever since. But one group that was in the lead from the start, and never seems to let go, are men. There is some imbalance between how many male and female prize winners there are, but how significant is this imbalance? And is it better or worse within specific prize categories like physics, medicine, literature, etc.?

```
# Calculating the proportion of female laureates per decade
prop_female_winners <- nobel %>%
  mutate(female_winner = sex == "Female",
         decade = year - (year %% 10)) %>%
  group_by(decade, category) %>%
  summarize(proportion = mean(female_winner, na.rm = TRUE))

# Plotting the proportion of female laureates per decade
options(repr.plot.width=7, repr.plot.height=4)

ggplot(prop_female_winners, aes(x = decade, y = proportion,
                              color = category, shape = category)) +
  geom_point(size=3, alpha=0.6) +
  scale_y_continuous(labels = scales::percent,
                    limits = 0:1, expand = c(0,0)) +
  labs(title = "Female Noble Prize Winners by Decade") +
  theme(plot.title = element_text(hjust = 0.5))
```



## 7. Repeat laureates

For most scientists/writers/activists a Nobel Prize would be the crowning achievement of a long career. But for some people, one is just not enough, and there are few that have gotten it more than once. Who are these lucky few?

```
nobel %>%
  count(full_name) %>%
  group_by(full_name) %>%
  filter(n > 1)
```

```
## # A tibble: 6 x 2
## # Groups:   full_name [6]
##   full_name                                     n
##   <chr>                                         <int>
## 1 Comité international de la Croix Rouge (International Committee of~ 3
## 2 Frederick Sanger                             2
## 3 John Bardeen                                 2
## 4 Linus Carl Pauling                           2
## 5 Marie Curie, née Skłodowska                  2
## 6 Office of the United Nations High Commissioner for Refugees (UNHCR) 2
```

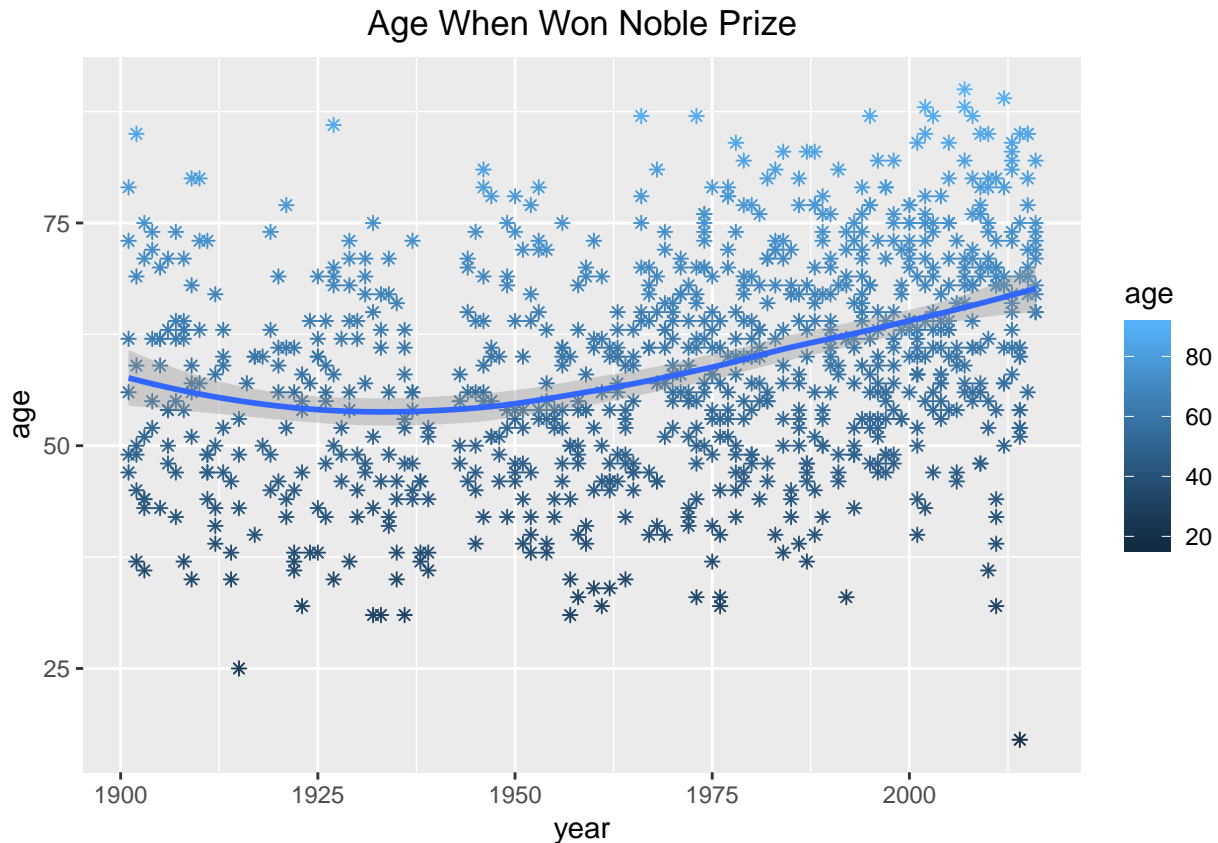
## 8. How old are you when you get the prize?

The list of repeat winners contains some illustrious names! We again meet Marie Curie, who got the prize in physics for discovering radiation and in chemistry for isolating radium and polonium. John Bardeen got it twice in physics for transistors and superconductivity, Frederick Sanger got it twice in chemistry, and Linus Carl Pauling got it first in chemistry and later in peace for his work in promoting nuclear disarmament. We also learn that organizations also get the prize as both the Red Cross and the UNHCR have gotten it twice.

But how old are you generally when you get the prize?

```
# Calculating the age of Nobel Prize winners
nobel_age <- nobel %>%
  drop_na(birth_date, year) %>%
  mutate(age = year - year(birth_date))

ggplot(nobel_age, aes(x = year, y = age, color = age)) +
  geom_point(shape=8) +
  geom_smooth() +
  labs(title = "Age When Won Noble Prize") +
  theme(plot.title = element_text(hjust = 0.5))
```



## 9. Age differences between prize categories

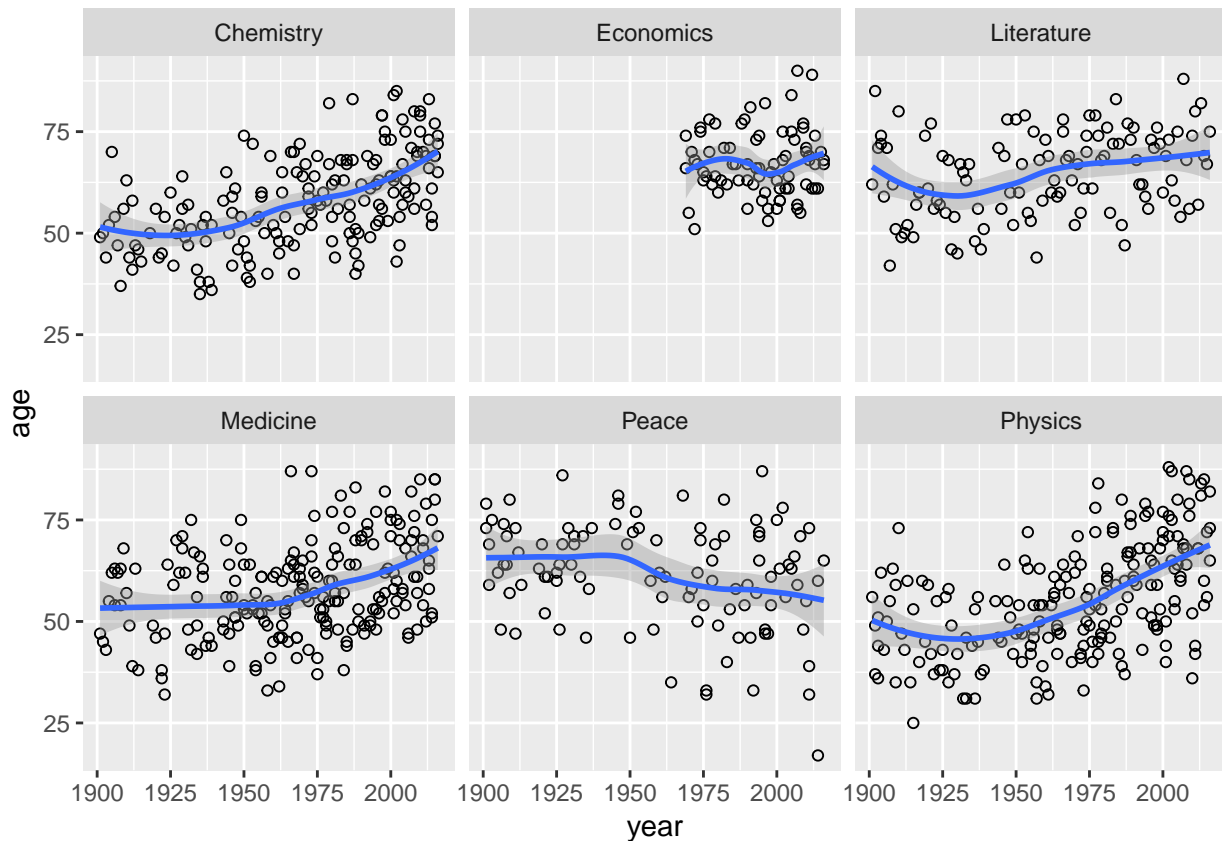
The plot above shows us a lot! We see that people used to be around 55 when they received the prize, but nowadays the average is closer to 65. But there is a large spread in the laureates' ages, and while most are 50+, some are very young.

We also see that the density of points is much higher nowadays than in the early 1900s – nowadays many more of the prizes are shared, and so there are many more winners. We also see that there was a disruption in awarded prizes around the Second World War (1939 - 1945).

Let's look at age trends within different prize categories.

```
# Same plot as above, but faceted by the category of the Nobel Prize
ggplot(nobel_age, aes(x = year, y = age)) +
  geom_point(shape = 1) +
  geom_smooth() +
  facet_wrap(~category)
```





## 10. Oldest and youngest winners

Another plot with lots of exciting stuff going on! We see that both winners of the chemistry, medicine, and physics prize have gotten older over time. The trend is strongest for physics: the average age used to be below 50, and now it's almost 70. Literature and economics are more stable, and we also see that economics is a newer category. But peace shows an opposite trend where winners are getting younger!

In the peace category we also a winner around 2010 that seems exceptionally young. This begs the questions, who are the oldest and youngest people ever to have won a Nobel Prize?

```
# The oldest winner of a Nobel Prize as of 2016
```

```
kable(nobel_age %>%
  top_n(1, age))
```

year	category	prize	motivation
2007	Economics	The Sveriges Riksbank Prize in Economic Sciences 2007	“for having laid the foundations of mechanis

```
# The youngest winner of a Nobel Prize as of 2016
```

```
kable(nobel_age %>%
  top_n(1, desc(age)))
```

year	category	prize	motivation
2014	Peace	The Nobel Peace Prize 2014	“for their struggle against the suppression of children and young people an

## 11. You get a prize!

Hey! You get a prize for making it to the very end of this notebook! It might not be a Nobel Prize, but I made it myself in paint so it should count for something. But don't despair, Leonid Hurwicz was 90 years old when he got his prize, so it might not be too late for you. Who knows.

The youngest winner ever who in 2014 got the prize for “[her] struggle against the suppression of children and young people and for the right of all children to education”?

```
# The name of the youngest winner of the Nobel Prize as of 2016  
youngest_winner <- "Malala Yousafzai"
```